NAMRL - 1250

LEVEL

12

CHARACTERIZING SPATIAL ABILITY: DIFFERENT MENTAL PROCESSES REFLECTED IN ACCURACY AND LATENCY SCORES

Dennis E. Egan

August 1978

NAVAL AEROSPACE MEDICAL RESEARCH LABORATORY
PENSACOLA FLORIDA

6

Approved for public release; distribution unlimited.

6 CHARACTERIZING SPATIAL ABILITY: DIFFERENT MENTAL PROCESSES REFLECTED IN ACCURACY AND LATENCY SCORES

10 Dennis E. Egan, Ph.D.

9 Interim rept.

12 36p.

Naval Air Systems Command
W 43-13 8881

Naval Medical Research and Development Command
ZF51.524.004-2011

14 NAMRL-1250

Approved by

Ashton Graybiel, M.D.
Assistant for Scientific Programs

Released by

Captain R. E. Mitchel, MC, USN
Commanding Officer

11 August 1978

16 F51524

17 ZF51524004

Naval Aerospace Medical Research Laboratory
Naval Air Station
Pensacola, Florida 32508

SUMMARY PAGE

THE PROBLEM

Recent experimental studies have analyzed the time to perform tasks patterned after standard tests of spatial ability. Based on these analyses, information-processing models have been developed suggesting that subjects work through a sequence of component mental processes (e.g., code, transform, match) to perform spatial test items. If these models are correct, then response latencies, especially estimates of component-process durations, may be the best measures of spatial ability. By contrast, traditional psychometric analyses of these tasks have consistently used overall accuracy scores as measures of spatial ability.

A model of the relationship between traditional accuracy measures of spatial ability and theoretically based latency measures is proposed. In this model overall accuracy and mean latency are viewed as composite scores consisting of the product (accuracy) or sum (latency) of component-process parameters. Three experiments investigated the relationship between spatial accuracy and latency scores, and established some psychometric properties (reliability, correlation across tests, predictive validity) of various measures.

FINDINGS

While accuracy and mean latency scores each proved to be reliable and consistent across different tests, the two measures were virtually independent. Further analyses using component-process latency scores suggest that different mental processes influence overall accuracy and mean latency. One hypothesis consistent with the data is that spatial accuracy scores reflect the ability to accurately code a pictorial stimulus, but mean latency scores on the same items reflect the ability to mentally transform the code. Implications for ability testing are discussed.

Current address of the author is: Dennis E. Egan, Ph.D., Bell Laboratories, 600 Mountain Avenue, Murray Hill, New Jersey 07974.

INTRODUCTION

Recent experimental studies have analyzed human performance on tasks adapted from standard tests of mental abilities. In these new analyses tasks are broken down into a sequence of component processes, and response time is shown to consist of the sum of process latencies. Processing models based on response latency have been developed for spatial visualization (14), letter series completion (12), verbal and geometric analogies (16), and mental arithmetic (9).

If examinees perform test items by working through a sequence of processes in real time, then response latencies, especially estimates of component-process latencies, may be the best measures of the abilities being tested. In contrast, traditional psychometric analyses of these tasks have consistently used overall accuracy scores as measures of subjects' ability. Spatial ability serves as a case in point.

PSYCHOMETRIC ANALYSES OF SPATIAL ABILITY

Until recently, spatial ability had been studied exclusively by using factor analyses of accuracy scores from batteries of paper-and-pencil, multiple-choice tests. Kelly (11) and Thurstone (18) were among the first to induce the existence of a "spatial factor." Since then, efforts have concentrated on isolating two or more spatial factors through refinements in testing and statistical procedures (4, 7). Guilford's (6) identification of three spatial factors, cognition of visual-figural systems (CFS-V), cognition of figural transformations (CFT), and cognition of kinesthetic-figural systems (CFS-K), represents a current view of the factor structure of spatial ability. Besides being valid predictors of success in mechanical and technical training programs, tests loading on the first two of these factors have also proved to be valid predictors of pilot and navigator training criteria (7).

INFORMATION-PROCESSING ANALYSES OF SPATIAL VISUALIZATION

Performance on tests loading on the CFT factor has been studied recently by using latency of response to analyze the mental processing of individual items. Shepard and Metzler (14) studied a task in which pictures of two three-dimensional block structures were presented and subjects had to decide whether the two figures were the same or different. The main finding was that the latency to make a correct "same" response was linearly related to the angle through which one figure had to be mentally rotated to bring it into congruity with the other figure. Just and Carpenter (10) showed that eye movements and response latencies suggest that subjects work through a sequence of three processes - search, transformation and confirmation as they perform the block rotation task. In related tasks Shepard and Feng (13) found a linear relationship between the number of operations required for mental paper-folding items and the latency of response, and Cooper and Shepard (2) extended the findings to the mental rotation of individual letters.

A MODEL OF ACCURACY AND LATENCY MEASURES FOR COGNITIVE TASKS

How might subjects' characteristic accuracy and response latency on a cognitive task be related? Suppose subjects work through a sequence of processes to arrive at an answer to a test item. Figure 1 depicts such a sequence. For spatial tasks Process 1 might be coding the visual stimulus, Process 2 might be transforming the coded representation in some way (e.g., rotation), Process might be outputting the response by pushing a button, and so forth. While it is not the present intention to test a component-process model for spatial tasks, models have been developed elsewhere (3, 10) that follow this approach.

Given that subjects work through a sequence of processes, performance of Subject on Process is characterized by two parameters:

p_{ij} = P (Subject i completed Process j correctly), and

t_{ij} = Time taken for Subject i to complete Process j correctly.

These parameters cannot be estimated directly unless component processes are isolated (16, 17, 19). However, the following quantities are natural to consider because they are easily estimated and reasonably reliable:

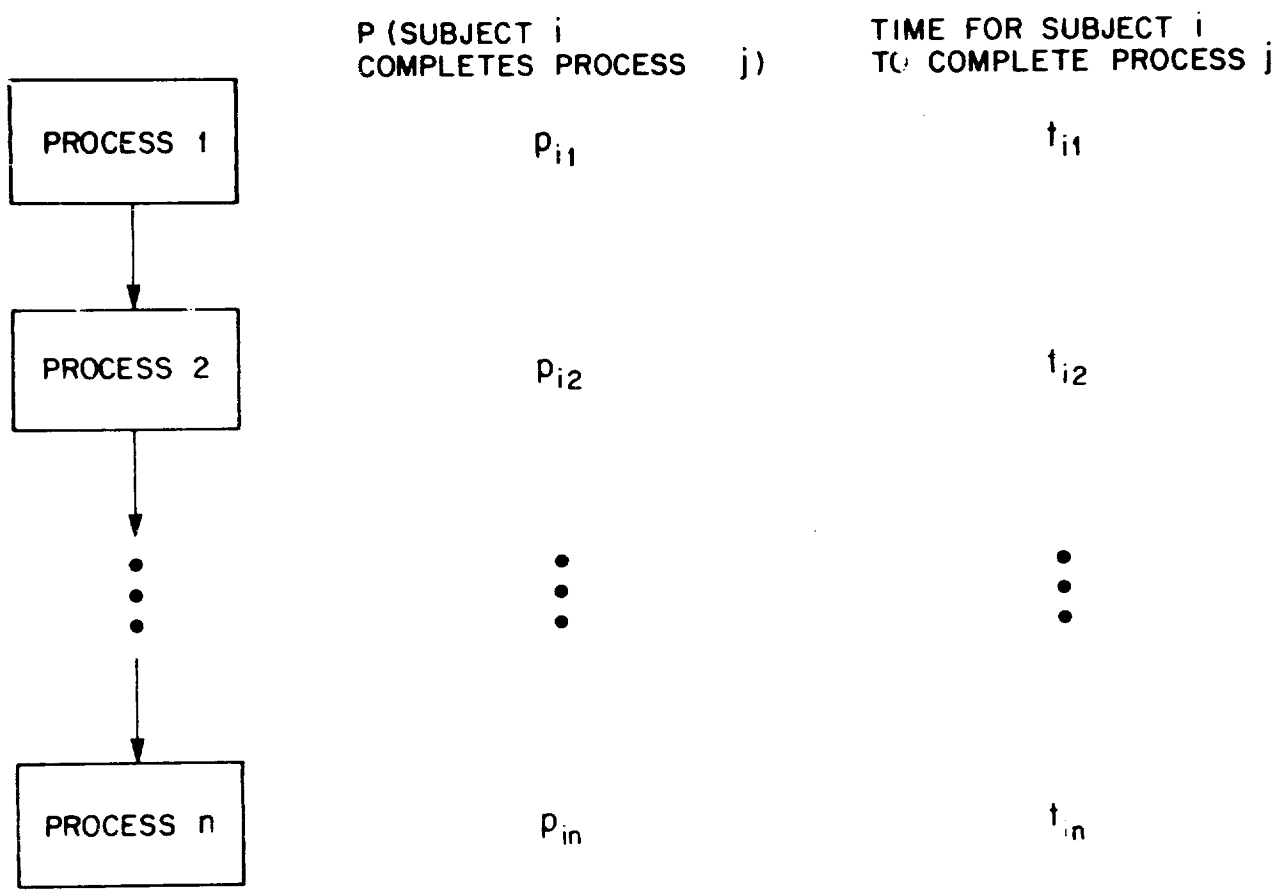
$p_{i.} = \prod_j p_{ij}$ = P (Subject i completes all processes correctly), and

$t_{i.} = \sum_j t_{ij}$ = Time taken for Subject i to complete all processes correctly.

If the test items are homogeneous, then the proportion of items a subject gets correct* is an estimate of $p_{i.}$, and the mean time taken to answer correctly is an

----

*Proportion correct and mean correct response latency are actually biased estimates of $p_{i.}$ and $t_{i.}$ because responses may result from guesses or a processing sequence in which more than one process is incorrect. For example, a sequence of two processes in a two-choice task may produce a correct response when both processes are completed correctly or when both processes are carried out incorrectly. While the bias will be ignored in this discussion, it may be quite important for tests composed of very difficult items.



$$p_{i.} = \prod_j p_{ij} = \text{P (SUBJECT } i \text{ COMPLETES ALL PROCESSES CORRECTLY)}$$

$$t_{i.} = \sum_j t_{ij} = \text{TIME FOR SUBJECT } i \text{ TO COMPLETE ALL PROCESSES CORRECTLY}$$

Figure 1. Accuracy and latency parameters of a component-process model and definitions of composite accuracy and latency scores.

estimate of $t_{i\cdot}$. The measure $p_{i\cdot}$ should be closely related to traditional accuracy scores, while $t_{i\cdot}$ is the sum of theoretical process latencies.

What relationship ought to exist between $p_{i\cdot}$ and $t_{i\cdot}$? It should be clear that $p_{i\cdot}$ and $t_{i\cdot}$ are both composite scores and that the character of these composites depends on the variability and intercorrelation of their components. Thus the correlation of the two composite scores may vary considerably, depending on which processes influence them.

Let us take the case of a sequence of two component processes for a task. Then for subject i the probability of completing both processes correctly is $p_{i\cdot} = p_{i_1} p_{i_2}$ and the time taken to complete both processes correctly is $t_{i\cdot} = t_{i_1} + t_{i_2}$. Let us further assume that $t_{i_1}, t_{i_2}, p_{i_1}$ and p_{i_2} have a multivariate normal distribution. This rather general model has 14 parameters for this example, that is, 4 means (t_1, t_2, p_1, p_2), 4 variances $s^2_{t_1}, s^2_{t_2}, s^2_{p_1}, s^2_{p_2}$, and $\binom{4}{2}=6$ correlations between variables. To make the model more tractable, we will add three further assumptions about the correlations. First, we will assume that $r_{t_{i_1}, t_{i_2}} = r_{p_{i_1}, p_{i_2}} = \rho_B$. That is, correlations between processes are equal for latency and accuracy measures. Second, we will assume that $r_{t_{i_1}, p_{i_1}} = r_{t_{i_2}, p_{i_2}} = \rho_W$. That is, correlations between latency and accuracy within processes are equal for both processes. Third, we will assume that once ρ_B and ρ_W are known, $r_{t_{i_1}, p_{i_2}}$ and $r_{t_{i_2}, p_{i_1}}$ add no further information. That is, the latency of one process correlates with the accuracy of the other only to the extent determined by ρ_B and ρ_W.

Given these assumptions, we have:

$$E(t_{i\cdot}) = \bar{t}_1 + \bar{t}_2$$

$$Var(t_{i\cdot}) = s^2_{t1} + s^2_{t2} + 2\rho_B s_{t1} s_{t2}$$

$$E(p_{i\cdot}) = \bar{p}_1\bar{p}_2 + \rho_B s_{p1} s_{p2}$$

$$Var(p_{i\cdot}) = \bar{p}_1^2 s^2_{p2} + \bar{p}_2^2 s^2_{p1} + s^2_{p1} s^2_{p2} (1 + \rho_B)^2 + 2\rho_B s_{p1} s_{p2} \bar{p}_1 \bar{p}_2$$

$$Cov(t_{i\cdot}, p_{i\cdot}) = \rho_w (\bar{p}_1 s_{p2} (\rho_B s_{t1} + s_{t2}) + \bar{p}_2 s_{p1} (\rho_B s_{t2} + s_{t1})).$$

Then R = Correlation between $t_{i\cdot}$ and $p_{i\cdot}$.

$$= \frac{Cov(t_{i\cdot}, p_{i\cdot})}{\sqrt{Var(t_{i\cdot})\, Var(p_{i\cdot})}}.$$

It is evident from the covariance formula that the underlying relationship of component accuracy and latency cannot be known by simply observing the correlation between $t_{i\cdot}$ and $p_{i\cdot}$. For example, a zero correlation of accuracy and latency within processes implies a zero correlation of composite (i.e., $\rho_w = 0$ implies $(R = 0)$, but the converse is not true. In fact ρ_w need not have the same sign as R, depending on ρ_B and parameter variances.

The point of this model is to show that accuracy and latency measures of component processes may be correlated even when accuracy and latency composites are virtually independent. Put another way, component accuracy and latency may be measuring the same ability even when composite accuracy and latency effectively measure different abilities. This can be due simply to the variance and intercorrelation of component-process parameters combined in composite scores. In particular, it may be the case that an accuracy composite such as $p_{i\cdot}$ is heavily influenced by accuracy in one process while a latency composite such as $t_{i\cdot}$ is heavily influenced by the latency of a second process uncorrelated or negatively correlated with the first.

OBJECTIVES OF PRESENT STUDIES

The present studies were designed to obtain empirically correlations between accuracy and latency scores on spatial tests. The accuracy scores were estimates of $p_{i\cdot}$. Component-process accuracy scores could not be estimated reliably in the short testing sessions used. The latency scores ranged from the mean time taken to answer items correctly (i.e., estimates of $t_{i\cdot}$) to component-process latencies (i.e., estimates of t_{ij}'s). The main purpose was to determine the relationship between the new latency measures of spatial ability and the traditional accuracy scores on the same items. Additionally, psychometric characteristics of accuracy and latency scores (reliability, correlations across different tests, validity for predicting spatially loaded aviation training criteria) were obtained.

Three studies were performed. In the first two studies, subjects were given standard group tests of spatial ability (as part of an aptitude battery) as well as tests using items whose content was similar, but whose design permitted the analysis of accuracy and latency of response to individual items. Accuracy and mean latency scores were correlated, and factors were extracted. In the third experiment, two tests of spatial ability and a simple test of nonspatial judgments were administered using a test-retest procedure. This permitted an analysis of reliability and partial correlation of several component-process latency scores as well as an additional assessment of the accuracy-mean latency relationship. As subjects in these experiments were naval aviation candidates, the validity of latency and accuracy measures in predicting success in training programs for pilots and flight officers was also explored.

EXPERIMENTS I AND II

In the first experiment were drawn items from tests loading on two spatial factors. Items from the Guilford-Zimmerman Aptitude Survey's* Spatial Orienta-

----

*Permission was obtained from Sheridan Psychological Services, Inc., to modify the Spatial Orientation and Spatial Visualization subtests of the Guilford-Zimmerman Aptitude Survey.

tion (GZO) subtest and the U.S. Navy's Spatial Apperception Test (SPA) were used to represent Guilford's CFS-V factor. Items from the Guilford-Zimmerman Aptitude Survey's Spatial Visualization subtest (GVZ) represented the CFT factor. In the second experiment a block rotation test whose standard forms also load on the CFT factor replaced the modified GZO items.

PROCEDURE

Test Construction

Spatial Apperception Test. The new version of the SPA designed for latency scoring (LSPA) was constructed from multiple-choice items from Form A and Form B of the SPA. The LSPA requires examinees to judge whether a landscape shown in one panel of a slide is the view that would be seen from the cockpit of an airplane shown in another panel. The standard SPA presents for each of 30 landscapes a set of five airplanes shown at different orientations. An item from each test is given in Figure 2.

In the SPA an examinee selects the best choice for each item and has a time limit of 10 minutes for the entire test. In the LSPA subjects had a maximum of 15 seconds per item to make a "Yes" or "No" response. The 60 items for the LSPA were interleaved in order from the two forms of the SPA. Half the items were randomly selected to be Yes items, and the other half were No items. For Yes items the landscape was matched with the correct airplane from the SPA. For No items, the landscape was paired with a randomly selected false choice.

Spatial Visualization Test. The LGZV (see Figure 2) was constructed in a similar manner from the 40-item multiple choice GZV (Form B). The GVZ requires examinees to mentally rotate an alarm clock in a specified sequence and then judge which of five figures matches its final position. In the GZV the examinee has a time limit of 10 minutes for the entire 40-item test. In the LGZV subjects were given a maximum of 20 seconds per item to make a Yes or No response. Items in the LGZV were presented in the same order as they occurred in the GZV.

Spatial Orientation Test. The LGZO was constructed from Form A of the GZO (see Figure 2). This test requires examinees to determine whether a symbol accurately portrays the change in position and direction that has occurred from the top to the bottom drawing of a motorboat heading toward a coastline. The time limit on the 60-item GZO is 10 minutes. In the LGZO, subjects were given a maximum of 15 seconds to respond Yes or No to each item. The order of presentation was the same in the two tests, and selection of true and false items in the LGZO was again determined randomly.

Block Rotation Test. For the block rotaton test (LBRT), three rigid three-dimensional block structures were drawn, similar to those used by Shepard

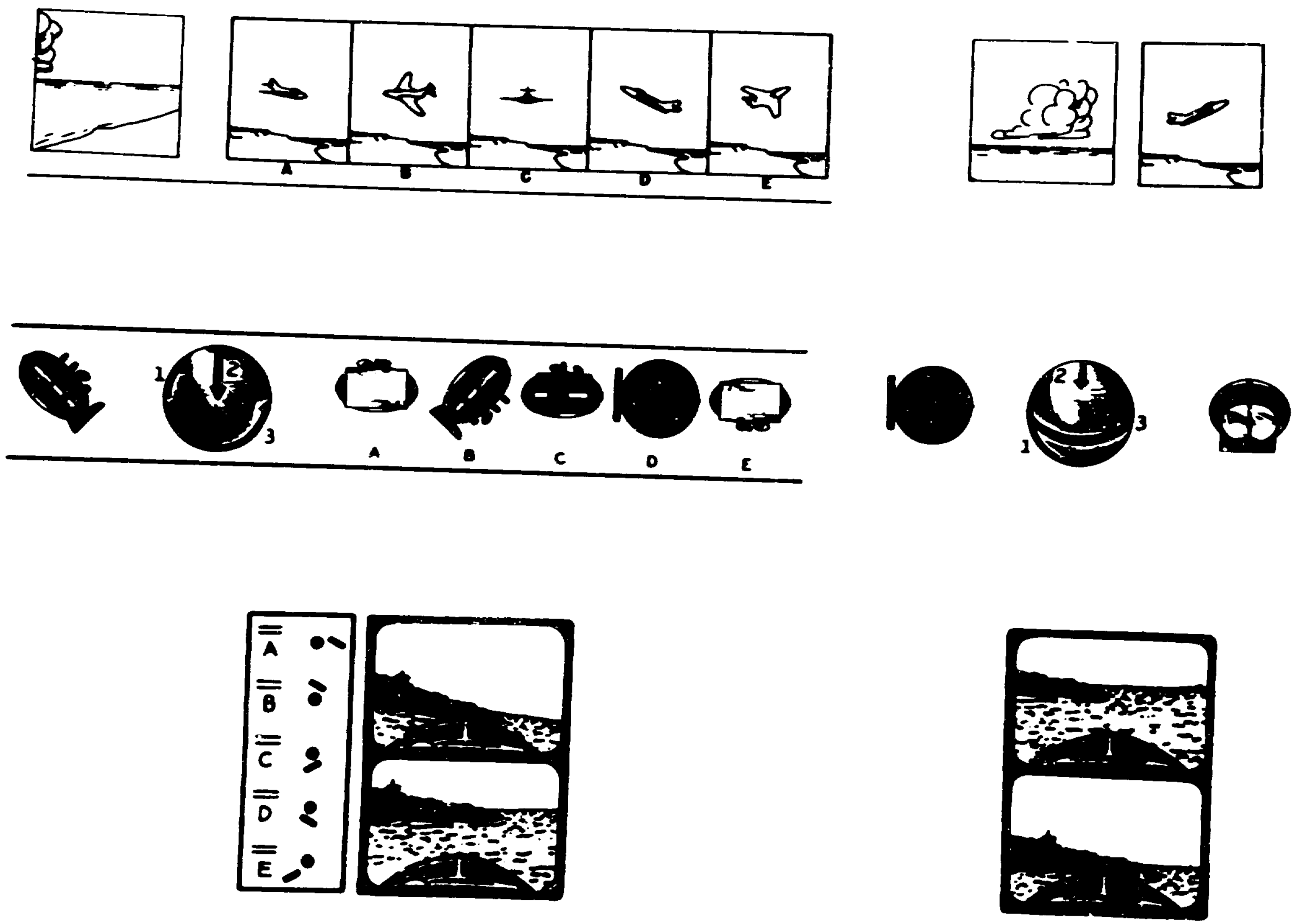


Figure 2. Examples of standard paper-and-pencil spatial test items and items redesigned for latency scoring: (top) SPA and LSPA, (middle) GZV and LGZV , (bottom) GZO and LGZO.

and Metzler (14). Photographs of the drawings and their mirror images were taken. Each test item consisted of either the same block structure presented in two different orientations, or a structure and its mirror image. Three sets of items were constructed, one for each block figure. In each set, 9 items presented a pair of identical figures at varying orientations, and 9 items presented a figure and its rotated mirror image. The nine match figures in each set differed by 0°, $\pm 40^{\circ}$, $\pm 80^{\circ}$, $\pm 120^{\circ}$, or $\pm 160^{\circ}$. The total number of items was 54, 9 match and 9 no-match items for each of three basic figures. The deadline for the LBRT was set at 20 seconds for each item.

Instructions. Instructions for the redesigned tests were simple modifications of the instructions for the paper-and-pencil forms. The modified instructions showed examples of the items and explained the use of the test apparatus. They also included a statement to be as accurate as possible, and informed the subjects of the maximum time allowed for each item.

Test Apparatus

The new tests were given on an automated system that controlled the presentation, timing, and scoring of two-choice test items for groups of six or fewer subjects. The system comprised six testing stations, a Kodak Ectagraphic self-focusing slide projector (Model AF-2), and a centrally located viewing screen. A UNIVAC 418 computer operating in a real-time mode controlled the system. Test stations were arranged in a row parallel to the screen and between the screen and projector. The screen was 4.2 meters in front of the stations. The row of stations was placed so that the viewing angle at the two outboard stations was no larger than 30°. The projected images of test stimuli subtended visual angles ranging from approximately 12° for the LGZO items to approximately [illegible]0° for the LBRT items. Each station was equipped with a hand-held switch box on which two response buttons were mounted. The lefthand button was labeled "No" and the righthand button was labeled "Yes." Subjects were instructed to hold the box with both hands and use their thumbs to activate the buttons.

Method

Subjects were given the new tests in the following order: LSPA, LGZV, LGZO (Experiment I) or LBRT (Experiment II). All subjects took the LSPA and, depending on their schedules and the availability of equipment, subsequently received the remaining tests. Three to five days after taking the new tests, subjects were given the Guilford-Zimmerman Aptitude Survey. In addition to the standard versions of the GZO and GZV, this survey includes subtests of Verbal Comprehension, General Reasoning, Numerical Operations, Perceptual Speed, and Mechanical Knowledge. These paper-and-pencil forms were administered under group testing conditions with approximately 25 examinees per group. The SPA had been given prior to admission to the program; so, those scores were obtained from the subject's records.

The procedure for the new tests started with the subjects reading the instruction booklet. When all indicated they understood the instructions the experimenter initiated the test. Items were projected onto the screen and subjects responded by pushing the appropriate button on the switch box. An item remained in view on the screen until either all subjects responded or the maximum time limit was reached. Approximately 1.5 seconds after the first of those events occurred, the slide projector advanced to the next item. After a succession of six such items, a blank slide was presented for the maximum time allowed. This served as a short rest. Just prior to the initiation of the next sequence of six items, a "Ready" signal was given.

Scoring

Latency of response to an item was defined as the interval between the onset of presentation and the response to the item. If a subject did not answer an item by the end of the time limit, the item was scored as wrong, with a latency equal to the time limit.

Subjects

The 134 male subjects were naval Aviation and Flight Officer Candidates at Pensacola Naval Air Station. Because of scheduling difficulties and equipment failures, complete data were available for only 31 subjects in Experiment I and 48 in Experiment II. Each subject had been selected for admission into his respective program on the basis of a battery of screening tests that included the SPA. Consequently, typical subjects in these studies had greater spatial ability than average male college graduates.

RESULTS

Psychometric Properties

The means, standard deviations, and reliabilities of accuracy and latency scores are given in Table I. These data cannot be taken as norms, since the absolute values of scores, especially latency scores, depend on the design and calibration of the test apparatus. However, these data do permit several useful observations.

Split-half reliabilities of latencies were generally high and usually exceeded the reliabilities of the corresponding accuracy scores on the new tests. Reliabilities of latencies approximated the levels of reliability typical of accuracy scores on the standard five-choice tests.

The proportion of items answered correctly was greater on the new versions of the tests, since the probability of guessing correctly was higher, and subjects at least attempted each problem. The standard versions of the GZO and

Table I

Means, Standard Deviations and Reliabilities of Accuracy and Latency Scores

Measure	N	Mean	S.D.	Reliability[a]
LSPA Number (Proportion) Correct	127	45.30 (.76)	6.65 (.11)	.75
LSPA Correct Latency (sec.)	127	6.02	1.32	.94
LGZV Number (Proportion) Correct	106	26.70 (.67)	4.97 (.12)	.69
LGZV Correct Latency (sec.)	106	10.43	2.25	.84
LGZO Number (Proportion) Correct	32	44.84 (.75)	8.04 (.13)	.93
LGZO Correct Latency (sec.)	32	7.05	1.16	.92
LBRT Number (Proportion) Correct	60	45.88 (.85)	4.78 (.09)	.65
LBRT Correct Latency (sec.)	60	6.39	1.25	.92
SPA Number (Proportion) Correct	133	19.74 (.66)	5.72 (.19)	.71[b]
GZV Number (Proportion) Correct	100	25.03 (.63)	8.16 (.20)	.91[c]
GZO Number (Proportion) Correct	100	30.27 (.50)	11.63 (.19)	.88[d]

[a]Reliabilities computed by split-half technique correcting for length of test unless otherwise noted.

[b]Uncorrected alternate-form reliability reported by Gannon (5).

[c]Split-half reliability reported in Guilford and Zimmerman (8).

[d]Reliability estimated by administering test in two separately timed, equivalent halves, intercorrelating the part scores, and applying the Spearman-Brown formula (8).

the GZV are speeded tests so that items occurring later in these tests may never be attempted. The lower reliability for the new accuracy scores may be due partly to the binary-choice format of these tests and partly to the restriction in range of ability sampled. Reliability data for the standard GZV and GZO for this population were not collected.

The LGZV was the most difficult of the new tests. If corrected for guessing, accuracy scores would be considerably below that of the others. The mean latency for correct responses to LGZV items was the highest, and the time limit was exceeded on a greater proportion of items from the LGZV (.061) than the LSPA (.034), LGZO (.017), or LBRT (.055).

Intercorrelations

Correlations among accuracy and latency scores are shown in Table II. The pattern of correlation suggests that accuracy and latency scores measure different facets of spatial ability. First, with few exceptions, correlations among accuracy scores on all the spatial tests were statistically significant and had a mean of $\bar{r}$ = .45. The highest correlations between accuracy scores occurred when accuracy on the standard and redesigned forms of the same test was compared. The correlations between the GZV and LGZV ($\bar{r}$ = .74) and between the GZO and LGZO ($\bar{r}$ = .72) are satisfactorily close to alternate-form reliability when restriction of range of ability in the sample is considered. The correlation of accuracy scores on the SPA and LSPA was lower ($\bar{r}$ = .45) but still highly significant. This lower correlation may have resulted because subjects were previously screened partly on the basis of SPA scores, or because SPA scores were obtained from two different forms of the test given many months before the experiment. The difference between the CFS-V and CFT factors is not present in the accuracy data as correlations among different tests of the same factor are no higher than correlations among tests of different factors. Generally, the accuracy data indicate that these tests are measuring a common process or ability.

A second characteristic of the data in Table II is that the measures of latency are highly correlated (r = .55). Thus the latency of correctly solving spatial problems was a consistent characteristic of a subject across all four of the new tests.

Third, correlations between latency and accuracy scores were generally negative and of low magnitude, having a mean of $\bar{r}$ = -.22. To the extent that a reliable relationship existed between accuracy and latency it was in the direction of the more accurate subjects responding faster. The largest correlations of this type involved accuracy scores on the GZO and GZV, two standard tests in which response latency partly determines how many items are attempted and thus directly influences accuracy scores.

Table II

Correlations of Spatial Accuracy and Latency Scores[a]

	2	3	4	5	6	7	8	9	10	11
1. LSPA Number Correct	-.14	.42**	-.23*	.39*	-.40*	.40**	-.11	.45**	.44**	.54**
2. LSPA Mean Correct Latency		-.07	.49**	-.14	.58**	-.01	.45**	-.11	-.21*	-.17
3. LGZV Number Correct			-.26**	.30	-.10	.45**	-.19	.37**	.74**	.56**
4. LGZV Mean Correct Latency				-.43*	.46**	-.25	.75**	-.21*	-.41**	-.42**
5. LGZO[b] Number Correct					-.27	- -	- -	.26	.46**	.72**
6. LGZO[b] Mean Correct Latency						- -	- -	.00	-.37*	-.31
7. LBRT[c] Number Correct							-.26	.13	.55**	.53**
8. LBRT[c] Mean Correct Latency								.01	-.45**	-.27
9. SPA									.36**	.35**
10. GZV										.59**
11. GZO										

[a]Cell Ns range from 31 to 127. Analyses using minimum Ns and separating the two experiments resulted in the same pattern of correlation.

[b]Experiment I only.

[c]Experiment II only.

* $P < .05$

** $P < .01$

Factor Analysis

A matrix of correlations among scores from the LSPA, LGZV, LBRT, and the Guilford-Zimmerman Aptitude Survey was analyzed by a principal-components procedure with varimax rotation. Each correlation in the matrix was based on a sample size of at least 44. The rotated factor loadings are given in Table III.

Two distinct factors, one with high loadings for accuracy the other for latency of response to spatial problems, emerged. A third group factor was also present with moderate to high loadings for Numerical Operations, Verbal Comprehension, and General Reasoning. In addition to high loadings for all spatial accuracy scores, the first factor showed moderate loadings for General Reasoning and Mechanical Knowledge, a typical pattern for a spatial ability factor. The only nonspatial test with a substantial loading on the spatial latency factor was Perceptual Speed, a highly speeded test requiring detailed pattern matching. This test had a moderate negative loading, indicating that subjects who answered more items correctly on the Perceptual Speed test tended to have lower mean latencies.

DISCUSSION

Mean latency of solving spatial problems was a reliable measure and correlated consistently across several tests of spatial ability. However, accuracy and latency of solving spatial problems defined distinct factors. Referring to the model, $t_{i.}$ and $p_{i.}$ for a given task did not correlate as highly as a set of $t_{i.}$'s or a set of $p_{i.}$'s obtained for different tasks.

The new accuracy scores correlated highly with the standard accuracy scores; so, they represent what has been traditionally called spatial ability. What do the latency scores represent? For example, being able to solve a spatial problem quickly has little to do with general intelligence, as indicated by the low loadings of Verbal Comprehension, General Reasoning, and Numerical Operations on the spatial latency factor. Perhaps spatial mean latency scores reflect merely the variability in the latency of decision and output processes common to a variety of tasks.

EXPERIMENT III

The third experiment was designed to determine whether mean latency scores obtained on the spatial tests have a "spatial" quality or whether they reflect variation in simple decision and output processes. Two types of analyses were carried out. First, subjects taking the spatial tests also took a test measuring simple decision time. This task was designed to include the decision and output processes involved in the spatial tasks, while omitting the spatial transformation necessary to make a decision. Second, estimates of component-process latencies were obtained.

Table III

Rotated Factor Loadings for Spatial Accuracy, Spatial Latency and Guilford-Zimmerman Aptitude Survey Scores

Measure	Factor I	Factor II	Factor III
LSPA Number Correct	.66	-.06	.10
LSPA Mean Correct Latency	.02	.70	-.25
LGZV Number Correct	.82	-.01	.13
LGZV Mean Correct Latency	-.26	.84	.04
LBRT Number Correct	.77	-.17	-.03
LBRT Mean Correct Latency	-.19	.87	.16
GZAS Verbal Comprehension	.40	.16	.46
GZAS General Reasoning	.57	-.08	.45
GZAS Numerical Operations	-.08	-.09	.85
GZAS Perceptual Speed	.30	-.47	.40
GZAS Spatial Orientation	.76	-.23	.10
GZAS Spatial Visualization	.78	-.33	.04
GZAS Mechanical Knowledge	.51	-.35	-.07
Proportion of Variance Accounted For	.37	.14	.10

Specifically, slopes and intercepts of linear response latency functions across sets of problems in the LGZV and LBRT were calculated for individual subjects. For the LGZV, a subject was assumed to carry out the following processing sequence (3): first, code the stimulus; second, transform the coded representation as indicated by the test item; third, decide whether the transformed representation matches the answer; fourth, output "Yes" if a match occurs, or "No" if a mismatch occurs. In this model only the transformation process is affected by the number of mental turns the item requires. The slope of the response latency function across classes of items requiring one, two, three, or four mental turns is thus an estimate of the amount of time taken for each additional mental transformation. The zero-intercept of that function is an estimate of time taken for all other processes - coding, decision, and output.

For the LBRT the model proposed by Just and Carpenter (10) for "same" trials was adopted. The model assumes that subjects work through a sequence of three processes: search, transformation, and confirmation. Angular disparity between two figures has the greatest effect on the transformation process. Thus, the slope of the response latency function across classes of items differing in orientation by 0^{o}, 40^{o}, 80^{o}, 120^{o}, or 160^{o} is an estimate of the additional time taken for mental transformation per 40^{o} increment in the angular disparity of the block structures. The zero-intercept is an estimate of the combined latency of those processes unaffected by angular orientation. This includes some portion of search and confirmation latency, and presumably all response output latency.

Since the slopes in these two models represent increments of time taken for additional spatial transformation, they should have a distinctly spatial character. Intercepts should represent combined latency for coding, decision, and output processes. A test-retest procedure was employed to observe the reliability of the various measures and any effects due to learning.

PROCEDURE

Test Construction

Spatial Tests. The LBRT was the same test as that used in Experiment II. The LGZV was a modified (50-item) version of the test used in the first two studies, where the additional items were simply second presentations of the most reliable original items.

Yes/No Decision Test. A test of simple Yes/No decisions (LYNT) consisted of 60 slides half of which projected the word YES the other half the word NO in large black letters on the screen. Subjects simply pressed the button corresponding to the stimulus word. Items were randomly ordered and grouped in blocks of six as in the other tests. Instructions for the LYNT were worded in a way similar to the instructions for the other tests.

Test Apparatus

A portable test controller was constructed, and in a new seating configuration five examinees sat in two staggered rows of test stations. The first row was 2.6 meters from the screen and the second row was 3.8 meters from the screen. This configuration resulted in a maximum viewing angle of 13^{o} at the two outboard stations. The visual angle of test items ranged from approximately 7^{o} for the LYNT words to 17^{o} for the LBRT figure pairs.

Method

Examinees were given the tests in the following order: LGZV, LYNT, and LBRT. One to three days after the first session, examinees returned for the second session in which the first session's procedure was repeated. Assignment of examinees to test stations on the second day was balanced under the constraint that no examinee should sit in the same place on both days. The procedure for the LBRT and LGZV was identical to that in the first two studies. For the LYNT, each item remained in view for 5 seconds before advancing.

Scoring

For each subject on the LGZV, the latency of each response was paired with the number of mental turns required by the item, and the regression line relating response latency to number of turns was obtained. The slope and zero-intercept of this least-squares function were used in addition to overall accuracy and mean latency. For the LBRT, latencies of correct "same" responses were paired with the angle of rotation required to bring the two block figures into congruity. Again, slopes and zero-intercepts of the least-squares regression lines were used as additional measures.

Subjects

Subjects were 50 naval aviation and flight officer candidates. Due to scheduling difficulties, only 41 examinees were available for the retest, so all analyses were restricted to those with complete data.

RESULTS

Group Data on Spatial Tasks

LGZV. Performance scores on the four classes of LGZV items were averaged across subjects, yielding the results shown in Figure 3. For both days, these group data show that response latency and error rates increased monotonically as a function of the number of mental turns required. An analysis of variance of the latency data indicated that there was a significant ($F(3,120) = 509.22$,

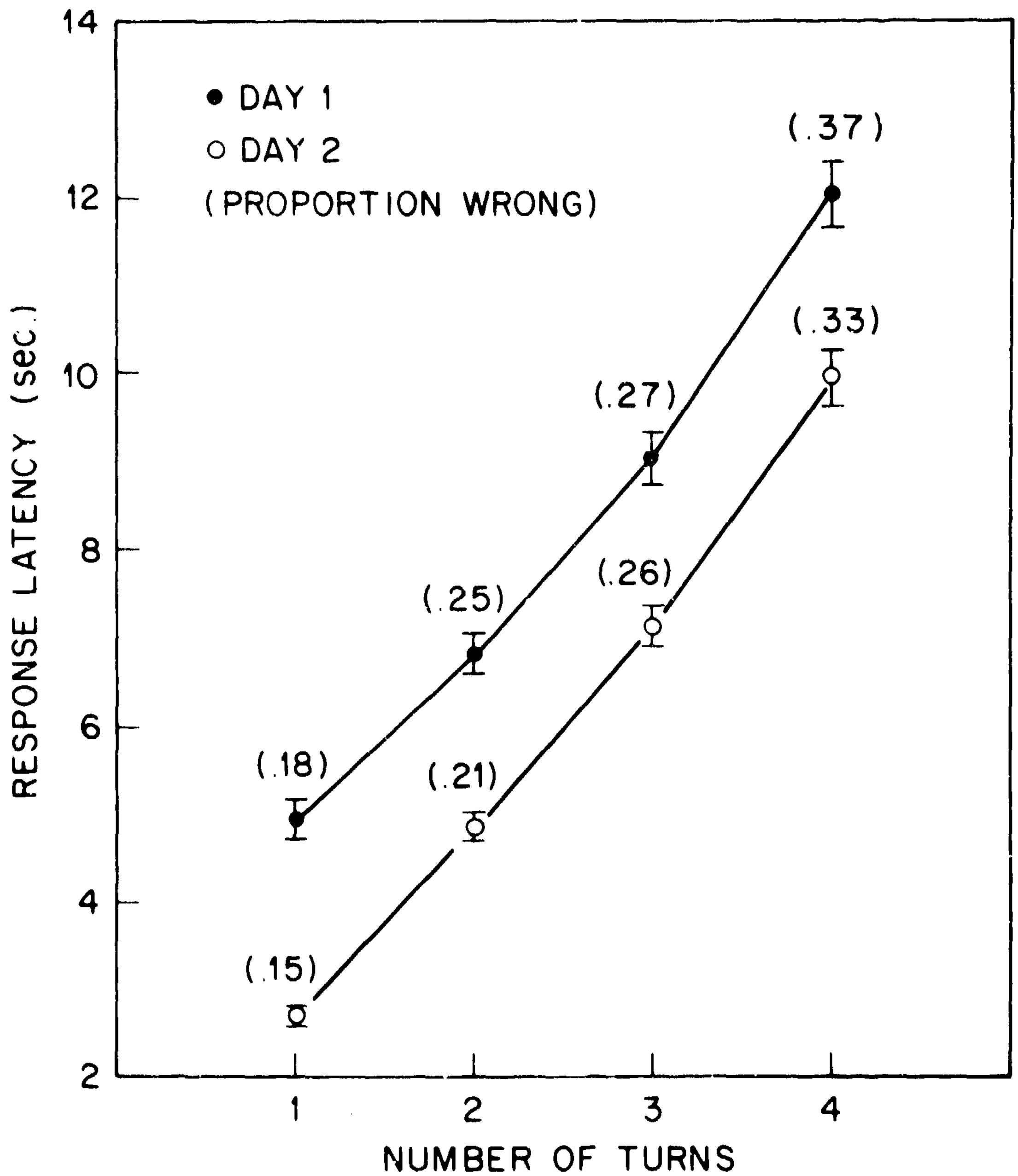


Figure 3. Latency and accuracy of responses to LGZV items grouped by the number of mental turns required. Numbers in parentheses represent the proportion of wrong answers for each class of item. Brackets are ± standard error of mean latency.

$p < .001$) increase in latency attributable to turns, a significant ($F(1,40) = 178.60$, $p < .001$) decrease in latency from Day 1 to Day 2, but no interaction of days and turns ($F(3,120) = 0.66$, $p > .10$). The linear trend across turns was highly significant ($F(1,40) = 662.15$, $p < .001$) and accounted for 99.3 percent of the variance due to turns. Consequently, there is ample evidence in the group data to support the use of slopes and intercepts to characterize subjects' latency of response

LBRT. For each subject on each day the mean response latency for correct "same" trials was computed for LBRT items requiring 40°, 80°, 120°, or 160° rotation. Scores for 0° rotation were excluded since they were based on only half the number of observations as other categories and proved to be unreliable. Scores from the other four categories were averaged across subjects with the results shown in Figure 4. The data show a monotonic trend such that items requiring greater rotation take longer to answer and result in more errors. An analysis of variance performed on the latency data showed that there was a significant ($F(3,120) = 107.99$, $p < .001$) increase in latency as more rotation was required, and a significant ($F(1,40) = 122.23$, $p < .001$) decrease in latency from Day 1 to Day 2, but only a small, nonsignificant interaction ($F(3,120) = 1.88$, $p > .10$). The linear trend across rotation was highly significant ($F(1,40) = 180.19$, $p < .001$) and accounted for 93.6 percent of the variance due to rotation. These group data are in agreement with the original Shepard and Metzler (14) findings and support the use of slope and intercept measures to characterize the response latency of individual subjects.

Psychometric Properties of Individual Measures

The means, standard deviations, and reliabilities of individual measures on Day 1 and Day 2 are given in Table IV. Subjects were more accurate and faster on Day 2, the effects being larger and statistically significant for the two spatial tests. As indicated in the group data for both the LGZV and LBRT, the time taken for coding, decision, and output processes as measured by intercepts decreased dramatically from Day 1 to Day 2. For the LBRT, there was also a significant decrease in slope from Day 1 to Day 2, meaning that a typical subject took less time for additional 40° increments in rotation on the second administration of the test.*

The reliabilities of spatial mean latency scores were lower than previously indicated. The discrepancy between split-half and test-retest reliabilities apparently reflects an instability of latency scores over time at least for subjects

---

*The apparent discrepancy between this finding and the lack of a Days x Rotation interaction in the group data is explained by the use of subjects' class means in the group analysis rather than individual data points which were used to compute slopes and intercepts for each subject.

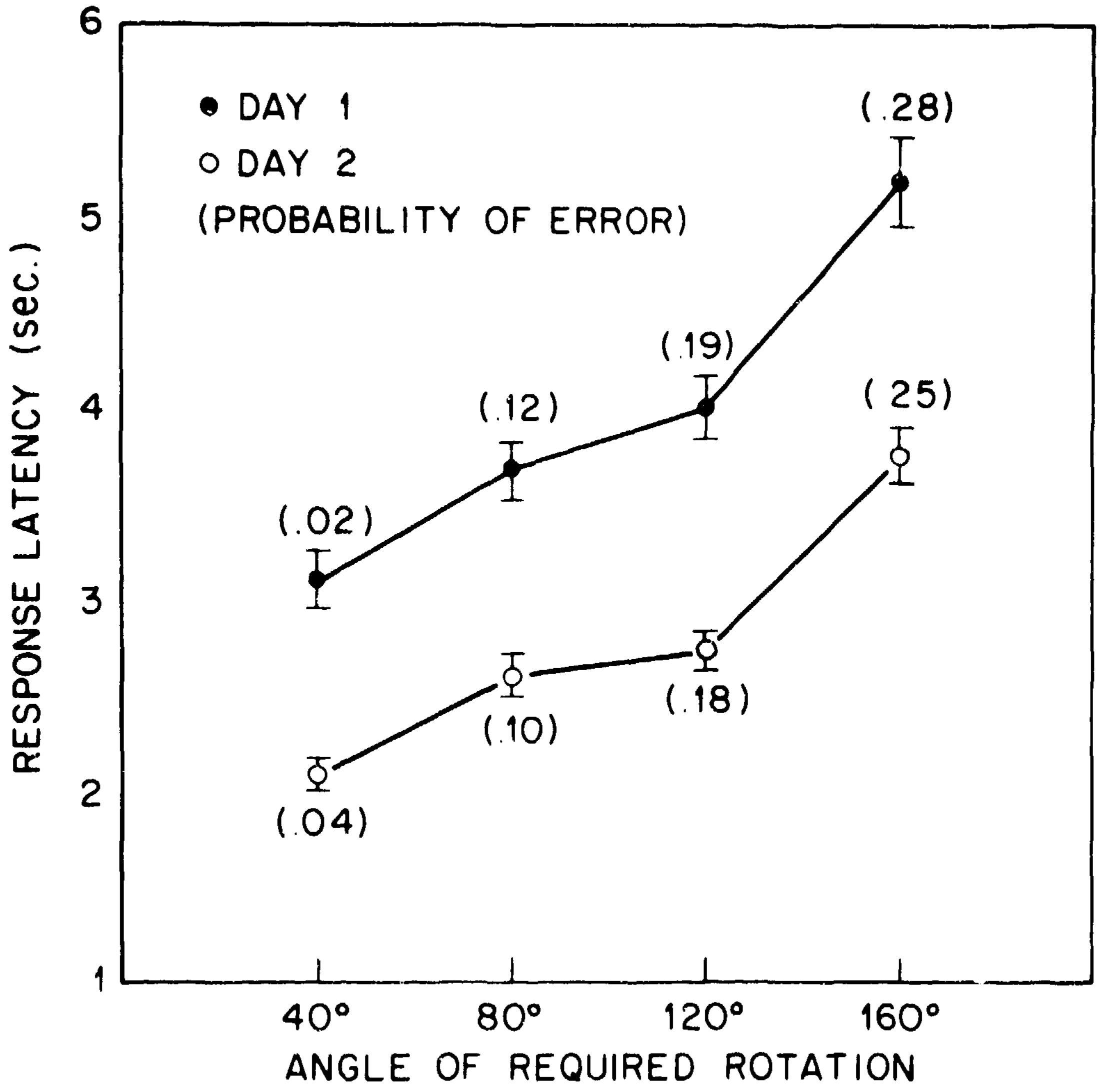


Figure 4. Latency of correct "same" responses and probability of error for LBRT items grouped by the angle of rotation required to bring the two figures into congruity. Numbers in parentheses represent the probability of error on each class of item. Brackets are + 1 standard error of mean latency.

Table IV

Means, Standard Deviations, Reliabilities, and t-values for

Measures on Day 1 and Day 2

Measure	Day 1 $\bar{X}$	Day 1 s.d.	Day 2 $\bar{X}$	Day 2 s.d.	r test-retest	t Day 1-Day 2
LGZV Number Correct[a]	43.34	5.23	44.95	6.31	.702**	-2.26*
LGZV Correct Latency (sec.)	7.80	1.48	4.89	1.06	.727**	12.01**
LGZV Least Squares Slope	2.34	.49	2.39	.35	.628**	-.52
LGZV Least Squares Zero Intercept	2.25	1.41	.11	.91	.562**	11.67**
LYNT Number Correct	52.32	1.46	52.73	1.30	.099	-1.43
LYNT Correct Latency (sec.)	.35	.07	.34	.05	.629**	.86
LBRT Number Correct	44.78	4.44	46.02	4.60	.669**	-2.16*
LBRT Correct Latency (sec.)	4.30	.88	3.05	.59	.705**	12.79**
LBRT Least Squares Slope	.64	.42	.50	.23	.414**	2.27*
LBRT Least Squares Zero Intercept	2.40	1.12	1.55	.67	.598**	6.09**

a *The worst case of exceeding the deadline was for the LGZV on Day 1 for which .018 of all responses were over the deadline.*

* $p < .05$

** $p < .01$

with little practice (1). The reliabilities of mean latency scores were still higher than reliabilities of corresponding accuracy scores for spatial tests. Slopes, intercepts, and mean latency on the LYNT had somewhat lower reliabilities. The reliability of accuracy scores on the LYNT was extremely low due to the fact that everyone had close to perfect accuracy on the task. This score was excluded from further analyses.

Intercorrelations

To improve the reliability of measures, each subject was assigned the mean of his Day 1 and Day 2 scores. These mean scores were correlated, and the results are given in Table V. The pattern found in the first two studies repeated. Spatial accuracy scores correlated significantly ($r = .53$, $p < .01$), and spatial mean latency scores correlated significantly ($r = .58$, $p < .01$). In this study accuracy and mean latency were virtually independent, the mean of accuracy-latency correlations being $r = .01$.

It might be hypothesized that the low accuracy-latency correlations are due to the mixture of items used. As shown in Figures 3 and 4, different classes of items on each spatial test resulted in different levels of performance. It is possible that high accuracy-latency correlations might exist within a class of items but are disguised by the averaging process necessary to obtain mean latency and overall accuracy. To check this hypothesis, accuracy and latency scores for individual subjects on each class of items in the LGZV and LBRT were obtained. The within-class correlations of accuracy and latency were no higher than the correlation of mean latency and overall accuracy. In fact, the within-class correlations of accuracy and latency were distinctly lower than correlations of accuracy scores across classes and correlations of latency scores across classes.

One question of interest is whether mean latencies from the LGZV and LBRT merely reflect decision and output processes. To resolve this question, the partial correlation between mean latency scores on the LGZV and LBRT was computed, holding the LYNT latency score constant. This partial correlation is $r = .54$ ($p < .01$) for these data, indicating that the relationship between mean spatial latencies cannot be due entirely to the latency of decision and output processes.

Further evidence of the "spatial" nature of the mean latency scores is derived from positive correlations between mean latencies and slopes. The LGZV slope correlated highly with both its own mean latency ($r = .70$, $p < .01$) and that of the LBRT ($r = .42$, $p < .01$). The LBRT slope also correlated positively but at lower levels with its mean latency ($r = .30$, $.10 > p > .05$) and the LGZV mean latency ($r = .16$, $p > .10$).

The two slope measures correlated positively ($r = .33$, $p < .05$) as did the two intercepts ($r = .35$, $< .05$). The strong negative correlations between slopes

Table V

Correlations Among Accuracy, Latency, Slope and Intercept Scores

Measure	2	3	4	5	6	7	8	9
1. LGZV Number Correct	.04	.25	-.36*	-.20	.53**	.02	.18	-.27
2. LGZV Mean Correct Latency		.70**	.22	.30	.15	.58**	.16	.28
3. LGZV Least Squares Slope			-.52**a	.13	.37*	.42**	.33*	.00
4. LGZV Least Squares Zero Intercept				.22	-.33*	.09	-.29	.35*
5. LYNT Mean Correct Latency					-.10	.28	-.03	.25
6. LBRT Number Correct						-.16	-.05	-.36
7. LBRT Mean Correct Latency							.30	.52**
8. LBRT Least Squares Slope								-.60**a
9. LBRT Least Squares Zero Intercept								

* $p < .05$

** $p < .01$

[a] Slopes and intercepts show an artificially high negative correlation because their estimates are not independent.

and intercepts of the same task shown in Table V may be inflated because the estimators for these parameters are not independent. A lower bound on the true magnitude of these correlations was obtained, using the correlations between slopes and intercepts estimated in different sessions. For the LGZV the average slope-intercept correlation across sessions was $\bar{r}$ = .29; for the LBRT it was r = -.27. In both cases, even these estimates suggest that the two component processes are negatively correlated. Finally, as expected, response latency to Yes/No items tended to correlate more strongly with intercepts which include decision and output latency than with slopes which measure spatial transformation latency.

Relationship of "Spatial Ability" to Component Processes

It is possible to ask which of the latency measures are more strongly related to accuracy scores (i.e., traditional "spatial ability"). The result is somewhat surprising. The mean of the four accuracy-slope correlations was low (r = .19) and in the unexpected direction. Subjects with slower rates of mental rotation tended to have higher accuracy scores. On the other hand, all accuracy-intercept correlations were at least marginally significant with a mean of $\bar{r}$ = -.30. Subjects who could rapidly code, decide, and output were more accurate. In every case the correlation between accuracy and intercepts was stronger than the correlation between accuracy and mean latency. Thus "spatial ability" as commonly defined by accuracy scores appears to have more in common with coding, decision, and output latency than it does with spatial transformation latency or mean latency.

ANALYSIS OF PREDICTIVE VALIDITY

PROCEDURE

The utility of the various measures of spatial ability in predicting real-world criteria was explored as follows. Training records of the subjects participating in the first two experiments and other preliminary studies were examined, and training criteria were correlated with the spatial test scores. The criteria were academic and performance scores for Navy pilot and flight officer candidates obtained at different stages of training up to 18 months after participating in these experiments. For pilots the criteria used were: (i) the overall grade from Aviation Officer Candidate School (AOCS) that included performance in mathematics, physics, and engineering courses; (ii) a one/zero criterion of pass (1) or fail (0) during flight training; and (iii) for those students who passed, the flight performance grade obtained in 13 instructional flights prior to soloing in a light aircraft. For flight officers, AOCS grades, two grades from basic school reflecting academic performance (engineering, navigation, technical training) and performance on training flights, and again a one/zero, pass/fail criterion in basic school were used.

RESULTS

Correlations between criteria and measures from the LGZV are given in Table VI. Although the relationships tend to be weak because of the imprecision of both tests and criteria, certain regularities are apparent. The accuracy score from the LGZV correlated positively with the criteria. This is consistent with the findings that aviation training programs require a great degree of "spatial ability" as traditionally defined (7). It is also consistent with the contention that the nature of conventional spatial ability scores is preserved in the accuracy scores obtained by the present procedure. Intercepts and to a lesser extent mean latencies showed a negative relationship to the criteria, subjects with longer latencies tending to perform worse. In particular, this was true for the intercept's correlations with AOCS grade and the pass/fail criterion for flight officers. The rate of spatial transformation as measured by slopes had only a very weak positive relationship to the criteria. Thus, for these criteria, accuracy scores and an estimate of the time to code, decide, and output were more successful predictors than were mean latency or rate of mental rotation. This general pattern held when the other spatial tests were used as predictors.

GENERAL DISCUSSION

PROCESSES REFLECTED IN MEASURES OF SPATIAL ABILITY

Relationships Between Latency and Accuracy

It is useful to distinguish among three kinds of latency and accuracy relationships. If stimulus conditions are varied and both response latency and accuracy are measured, latency and accuracy typically show a strong negative correlation across stimuli. This is true of the present data where, for example, items requiring more rotation in the LGZV or a larger single rotation in the LBRT take longer to answer correctly and produce more errors. In cases such as these, accuracy and latency are two dependent measures of the same effect. A second kind of latency-accuracy relationship involves manipulating instructions, payoffs, or deadlines within stimulus conditions. In this case a speed-accuracy tradeoff may be produced. Fast responses are the result of guessing or incomplete processing and are therefore more likely to be errors. The extent to which this occurred in the present studies is not certain. While wrong answers typically took longer than correct answers of the same type (3), a speed-accuracy tradeoff cannot be conclusively rejected without more complete data.

The latency-accuracy relationship of primary interest in the present studies differs from each of the previous types. In the present studies subjects were allowed to respond within a fairly comfortable deadline to a variety of items. Characteristic accuracy and latency for individual subjects were measured and correlated. The two measures provided reliable but distinct information about the subjects.

Table VI

Correlations of LGZV Measures with Pilot

and Flight Officer Training Criteria

Measure	Pilots[a]		
	AOCS Grade	Flight Training Pass/Fail	Pre-Solo Flight Grade
LGZV Number Correct	.24*	.13	.25
LGZV Mean Correct Latency	.15	-.01	-.01
LGZV Least Squares Slope	.03	.14	.09
LGZV Least Squares Zero Intercept	.01	-.19	-.13

Measure	Flight Officers[b]			
	AOCS Grade	Basic School Pass/Fail	Basic School Academic Grade	Basic School Flight Grade
LGZV Number Correct	.55**	.16	.42**	.16
LGZV Mean Correct Latency	-.19	-.29*	-.03	-.16
LGZV Least Squares Slope	.20	.13	.06	-.02
LGZV Least Squares Zero Intercept	-.34**	-.32**	-.04	-.11

* $p < .05$

** $p < .01$

[a] The number of pilots starting was 75. Of these, 61 passed flight training and received grades.

[b] The number of flight officers starting was 76. Of these, 39 passed basic school and received grades.

WHAT DO MEASURES OF SPATIAL ABILITY MEASURE?

Accuracy scores on the redesigned tests (i.e., estimates of $p_{i \cdot}$) are closely related to traditional measures of "spatial ability." This is borne out by high correlations between accuracy on the standard and redesigned forms of tests, and by the validity of the new accuracy scores in predicting aviation training criteria. In terms of the model, the process (or processes) influencing traditional accuracy scores also contributes a large part of the variance to the new accuracy scores. Highly accurate performance on this process (or processes) has come to be called high "spatial ability."

As the results show, traditional "spatial ability" and the mean latency of response to the same test items loaded on different factors. More specifically, spatial ability as conventionally defined was only weakly related to spatial transformation latency. On the other hand, "spatial ability" correlated significantly with coding, decision, and output latency. One explanation of these results is that the accuracy score reflects variability mainly in the accuracy of coding or searching (10) pictorial stimuli, but the mean latency score reflects substantial variability in the speed of mentally transforming the code.

The hypothesis that "spatial ability" (i.e., accuracy) reflects variation in coding while mean latency reflects variation in rate of transformation is consistent with the following findings. 1) Accuracy and mean latency of response to spatial problems loaded on different factors. This result is expected if (a) the accuracy composite is influenced by variability in one process (e.g., coding), (b) the latency composite is influenced by variability in another process (e.g., transformation), and (c) the two processes are negatively correlated in the population. The latter condition is true and logically must be true in order to account for the significant negative correlation of intercepts with accuracy when mean latency and accuracy are virtually independent. 2) Mean latency correlated positively with slopes that measure rate of mental transformation. This is direct evidence of the nature of the $t_{i \cdot}$ composite for spatial tasks. 3) Spatial accuracy scores correlated negatively with intercepts. This result follows if the $p_{i \cdot}$ composites are influenced mainly by a process also measured by intercepts. The coding process is the most likely candidate. If accuracy were instead a measure of decision and output processes, then accuracy should have correlated higher with the LYNT latency. 4) Practice both improved accuracy and lowered intercepts. If practice or familiarity with stimuli improves a single process such as coding, then this result follows. 5) Accuracy and intercepts were the best predictors of aviation training criteria. This suggests that the aspect of spatial accuracy scores important for predictive purposes is also measured by intercepts. Of the processes assumed to be measured by intercepts, coding has the most "spatial" character.

An alternative explanation of these data would begin with the premise that there need be no correlation between accuracy and latency measures of a given process or task. In terms of the model the assumption would be that ρ_w equals zero so that R, the correlation of composite accuracy and latency, must equal zero. Latency would be construed as measuring a different "trait" (e.g., caution or motivation) totally divorced from the content of a test and perhaps only of minor interest. While this explanation has no difficulty with the near-zero correlation between accuracy and mean latency, it cannot easily explain several other aspects of these data. First, under the simplest such theory there is no reason why intercepts should correlate more strongly with accuracy than mean latency does. Second, as the results of Experiment III show, practice reduced errors and lowered the latencies of some components. It seems likely that different subjects would come to a test with the equivalent of different amounts of practice and that this should result in significant negative values of ρ_w. Third, if latency is a measure of a general trait, then correlations between Yes/No latency and spatial latencies should be as high as the correlation between two spatial latencies, which is not the case. Rather than measurement of different traits leading to low accuracy-latency correlations, the view espoused here is that differential contribution of processes leads to low correlation of composites.

The available data suggest that ρ_B and ρ_w have negative values for spatial rotation tasks. For ρ_B this claim is based on the negative correlations observed between slopes and intercepts. Evidence for negative ρ_w is less direct, but is suggested by the significant negative correlation of accuracy and intercepts. While these claims are consistent with all the findings, the model of latency and accuracy proposed here has not been rigorously tested. In further work the model's assumptions and the effects of violating those assumptions must be assessed. To do that will require enough data to reliably estimate component-process accuracy as well as latency. Minimally, any model of accuracy and latency scores for spatial tasks must deal with the fact that accuracy and mean latency may be virtually independent even when accuracy and the latency of a component process correlate significantly.

IMPLICATIONS FOR ABILITY TESTING

Since mean latency and accuracy measure distinct factors of spatial ability, the practice of using standard speeded tests is called into question. If much of the accuracy variance arises from one process, and much of the latency variance from another, then a speeded test makes sense only if the two processes correlate positively $(\rho_B > 0)$ and if accuracy and latency correlate negatively within processes $(\rho_w < 0)$. Of course, this need not be the case.

On the other hand, the investigation of component-process latencies is a promising direction for future research on cognitive abilities. Although component-process latencies were somewhat less reliable than mean latencies and overall accuracy, their reliability could be improved using longer testing sessions (see 1, 15). Latency measures of component processes almost certainly will be more reliable than corresponding accuracy measures. In the present studies, component-process latencies correlated consistently across different tests and to some extent were predictive of aviation training criteria.

Perhaps more important than psychometric considerations is the possibility that latency measures of component processes may lead to a greater understanding of spatial ability and other cognitive abilities. As suggested by the present results, characterizing subjects by reliable, factorially distinct, but rather gross measures (e.g., accuracy and mean latency) may not be optimal for some purposes. Such measures when viewed as composite scores may actually disguise important relationships between distinct component processes of a task and between two measures of the same process. Alternatively, use of component-process measures may provide more precise estimates of an ability and may aid in clarifying conceptual relationships among tasks.

REFERENCES

1. Chiang, A., and Atkinson, R. C., Individual differences and interrelationships among a select set of cognitive skills. Memory and Cognition, 4: 661-672, 1976.

2. Cooper, L. A., and Shepard, R. N., Chronometric studies of the rotation of mental images. In: W. G. Chase (Ed.), Visual Information Processing, New York: Academic Press, 1973.

3. Egan, D. E., Accuracy and latency scores as measures of spatial information processing. Pensacola, Florida: Naval Aerospace Medical Research Laboratory, February 1976.

4. French, J. W., The description of aptitude and achievement tests in terms of rotated factors. Psychometric Monographs, No. 5, 1951.

5. Gannon, F. B., Final Report for Contract Number Nonr-4688(00). Princeton, New Jersey: Educational Testing Service, March, 1969.

6. Guilford, J. P., The Nature of Human Intelligence. New York: McGraw-Hill, 1969.

7. Guilford, J. P., and Lacey, J. I. (Eds.), Printed classification tests. Research Report No. 5, Army Air Forces Aviation Psychology Program, 1947.

8. Guilford, J. P., and Zimmerman, W. S. The Guilford-Zimmerman Aptitude Survey. J. Appl. Psychol., 32: 24-34, 1948.

9. Hunt, E., Lunneborg, C., and Lewis, J., What does it mean to be high verbal? Cognitive Psychol., 7: 194-227, 1975.

10. Just, M. A., and Carpenter, P. A., Eye fixations and cognitive processes. Cognitive Psychol., 8: 441-480, 1976.

11. Kelly, T. L., Crossroads in the Mind of Man: A Study of Differentiable Mental Abilities. Stanford, Calif.: Stanford University Press, 1928.

12. Kotovsky, K., and Simon, H. A., Empirical tests of a theory of human acquisition of concepts for sequential events. Cognitive Psychol., 4: 399-424, 1973.

13. Shepard, R. N., and Feng, C., A chronometric study of mental paper folding. Cognitive Psychol., 3: 228-243, 1972.

14. Shepard, R. N., and Metzler, J., Mental rotation of three-dimensional objects. Science, 171: 701-703, 1971.

15. Snyder, C. R., Individual differences in imagery and thought. Unpublished doctoral dissertation. University of Oregon, June, 1972.

16. Sternberg, R. J., Intelligence, Information Processing, and Analogical Reasoning: The Componental Analysis of Human Abilities. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

17. Sternberg, S., The discovery of processing stages: Extensions of Donder's method. In: W. G. Koster (Ed.), Attention and Performance II. Acta Psychol, 30: 276-315, 1969.

18. Thurstone, L. L., Primary mental abilities. Psychometric Monographs, No. 1., 1938.

19. Ward, W. C., Individual differences in information processing units. Research Bulletin No. RB-73-70, Princeton, New Jersey, Educational Testing Service, December, 1973.

US GOVERNMENT PRINTING OFFICE: 1978-740-263/1534
Region No. 4

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NAMRL - 1250	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Characterizing Spatial Avility: Different Mental Processes Reflected in Accuracy and Latency Scores		5. TYPE OF REPORT & PERIOD COVERED Interim 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Dennis E. Egan, PhD.		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Aerospace Medical Research Laboratory Naval Air Station Pensacola, Florida 32508		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NAVAIR W43-13.8881 BUMED ZF51.524.004-2011
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Medical Research and Development Command National Naval Medical Center Bethesda, Maryland 20034		12. REPORT DATE August 1978 13. NUMBER OF PAGES Thirty Three (33)
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Human Performance; Mental Abilities; Process Latencies; Spatial Visualization; Information Processing; Latency and Accuracy Scores; Ability Testing		

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Recent experimental studies have analyzed the time to perform tasks patterned after standard tests of spatial ability. Based on these analyses, information-processing models have been developed suggesting that subjects work through a sequence of component mental processes (e.g., code, transform, match) to perform spatial test items. If these models are correct, then response latencies, especially estimates of component-process durations, may be the best measures of spatial ability. By contrast, traditional psychometric analyses of these tasks have consistently used overall accuracy scores as measures of spatial ability.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20.

A model of the relationship between traditional accuracy measures of spatial ability and theoretically based latency measures is proposed. In this model overall accuracy and mean latency are viewed as composite scores consisting of the product (accuracy) or sum (latency) of component-process parameters. Three experiments investigated the relationship between spatial accuracy and latency scores, and established some psychometric properties (reliability, correlation across tests, predictive validity) of various measures.

While accuracy and mean latency scores each proved to be reliable and consistent across different tests, the two measures were virtually independent. Further analyses using component-process latency scores suggest that different mental processes influence overall accuracy and mean latency. One hypothesis consistent with the data is that spatial accuracy scores reflect the ability to accurately code a pictorial stimulus, but mean latency scores on the same items reflect the ability to mentally transform the code. Implications for ability testing are discussed.